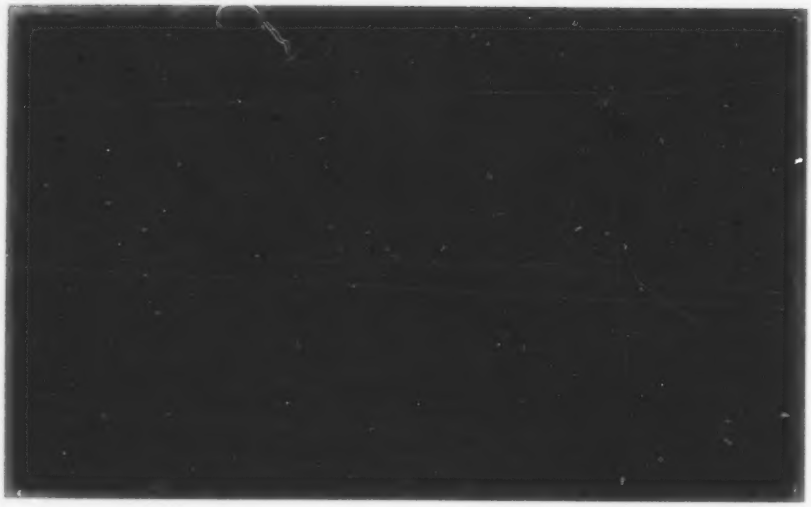
CENTRE DE RECHERCHE SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



TRAVAIL D'ÉTUDIANT

CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A.
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: 05-24146 BIBPOLYTEC
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

CRT-858

The Heuristic Identification of Urban Functional Land Use: Application to Shanghai

by

Jean-Paul Rodrigue

Département de Géographie et Centre de recherche sur les transports, Université de Montréal, C.P. 6128, succursale A, Montréal, Canada H3C 3J7.

Centre de recherche sur les transports — Publication #858

October 1992

Dépôt légal – Bibliothèque nationale du Québec,
Bibliothèque nationale du Canada, 1992

93-01790/

The Heuristic Identification of Urban Functional Land Use: Application to Shanghai[1]

Jean-Paul Rodrigue[2]
Department of geography
Université de Montréal
C.P. 6128, succ. A
Montréal, Québec
Canada, H3C 3J7

Abstract: Contemporary urban areas present an intense mixture of several types of land use and are extending over a vast territory. Over the past, the land use for a spatial entity was classified following a set of criteria. However, the association of a spatial entity to a land use type is not absolute, but must rather be viewed as the overlapping of diverse types of transportation infrastructures, movements, economic activities and population, each having its own type of land use. Considering this, a methodology to identify the functional land uses of spatial entities is developed. The approach is based on the assumption that a spatial entity has a membership value (or probability) to be associated with a type of land use. The assignation of a membership value is done with an heuristic classification procedure using elements of fuzzy logic. We advocate that land use can be conceived as knowledge about an urban system, and we develop a simple semantic network that represents a knowledge domain. The methodology helps the semantic network, a well known technique of knowledge based computing, "learn" about an urban system by finding the links between spatial entities and types of land use. The complexity of the urban system of Shanghai is used to test that new methodology in land use planning.
Key Words: Land Use Planning, Fuzzy Logic, Heuristics, Semantic Networks, Knowledge-Based Computing, Shanghai.

Résumé: Les aires urbaines contemporaines présentent un mélange intense de plusieurs types d'utilisation du sol et couvrent un vaste territoire. Initialement, l'utilisation du sol était classifiée selon un ensemble de critères. Cependant, l'association d'une entité spatiale à un type d'utilisation du sol n'est pas absolue, mais doit d'abord être vu comme la superposition de divers types d'infrastructures de transport, de mouvements, d'activités économiques et de populations, chacune ayant son propre type d'utilisation du sol. Considérant ceci, une méthodologie pour identifier l'utilisation fonctionnelle du sol d'unités spatiales est développée. L'approche est basée sur la supposition qu'une unité spatiale a une valeur d'appartenance (ou probabilité) d'être associée à un type d'utilisation du sol. L'assignation d'une valeur d'appartenance est faite avec une procédure de classification heuristique utilisant la logique des ensembles flous. Nous avancons que l'utilisation du sol peut se concevoir en tant que connaissance concernant un système urbain, et nous développons un réseau sémantique simple qui représente un domaine de connaissance. La méthodologie aide le réseau sémantique, une technique de computation à base de connaissance, à "apprendre" sur un système urbain en trouvant les liens entre les unités spatiale et les types d'utilisation du sol. La complexité du système urbain shanghaien est utilisée pour tester cette nouvelle méthodologie en planification d'utilisation du sol.
Mots Clés: Planification de l'utilisation du sol, Logique des ensembles flous, Heuristique, Réseaux sémantiques, Computation à base de connaissance, Shanghai.

1 Problem and Definition

One of the main difficulties of urban planning is the lack of information about a specific problem. The unavailability of information causes problems in the accurate representation of an urban system and can lead to mistaken planning decisions. With the scarcity of financial resources and available space and with the rising congestion levels of the transportation system, urban governments can ill-afford planning errors. So, methods to assess the state of an urban system are needed. One method of representing an urban system is to split it in several spatial entities, characterized by several types of land use. This favours a comprehensive gathering of statistical data. However, the level of political spatial aggregation of the urban system creates a problem of identification of land use. There is no clear membership of a spatial entity with a type of land use, even if this spatial entity is small in size. This is reinforced by the fact that all urban areas are an intense mixture of different types of land use including

---

[1] Paper presented at the International Conference on Transportation and Land Use, Shanghai, China, 1992.

[2] The author would like to thank Claude Comtois for usefull comments.

non urban activities and are extending over a vast territory. More specifically, multistoried building characterized by different activities at different levels represent several types of land use at one location. The smaller the scale of analysis, the more inaccurate is the identification of land use because the larger spatial entities are, the higher is the probability that they contain several types of land use. This have created problems of over simplification of the dynamics of an urban system that have not helped the accuracy of urban modelling and planning. From one side, planning requires the simplification of an urban system, but from the other, over simplification affects the efficiency of planning. It is possible to draw predominant types of land use for a spatial entity by using land use maps. However, the classification used does not necessarily represents the desired information nor the desired scale of detail. Criterias for classification of land use vary over time and from different sources. For example, the planning agencies of two cities may have different classification procedures of land use that can lead to quite different land use maps. The result is that the land use of different urban areas cannot be easily compared and analyzed. From another point of view, land use maps are not always available and agencies often lack the resources to periodically update the available maps and urban database. The design of land use maps often requires expensive field surveys using census data, areal photographs and questionnaires. It is still possible for small urban centres, but for large metropolitan areas the task is most of the time unaffordable. However, recent methodologies in remote sensing enable the production of land use maps (Lindgren, 1985). Those methodologies are not accurate in the sense that they are unable to classify specific types of urban land use like commercial, industrial and residential but offer a quick way to produce general land use maps..

Land use is a qualitative characteristic of space and the result of a dynamic process involving urban changes and spatial accumulation. In the literature, land use has been classified in two distinct categories: the formal and the functional land use (Mather, 1986). The formal land use is represented by qualitative characteristics such as the urban form, pattern and geometry. The functional land use represents the level of spatial accumulation in an urban area of movements, transportation infrastructures, economic activities and population. These concepts can be expressed by several quantitative variables. This article will look upon the functional land use. The later is more convenient to a dynamic representation of urban space because it shows the state of spatial accumulation of the variables at a given point in time, and therefore can be used as a planning tool such as modelling. Several transportation/land use models provide quantitative results, but it is to the user to deduce a functional land use with those results (Webster, Bly and Paulley, 1988). With those models, planners have a good idea of the socioeconomic consequences of urban planning, but the spatial consequences are disregarded. In transportation/land use modelling, the emphasis is far more on the spatial interaction aspect than the land use. An error often made is to consider a variable of territorial accumulation as land use. Population density cannot be considered as residential land use nor industrial output as industrial land use because those variables are not necessarily displaying characteristics which clearly relate to an urban land use pattern and density. We advance that those variables can be used to express a probability of existence of a specific type of land use. The better the knowledge about the relationships of spatial accumulation with land use, the more one or several variables can be linked with the presence of a land use type.

Recent methodological studies have stressed the need to integrate transportation/land use modelling in more geographical and graphical environment (Grothe and Scholten, 1990; van Est and Sliepen, 1990). It has been advanced that transportation/land use modelling and geographical information systems (GIS) could provide a powerful planning tool and research environment. Consequently, the first objective of the paper is to elaborate a methodology that can be used to identify with related variables, the most probable functional land uses for the spatial entities of an urban system in the form of a computer program. This provides a module that could be integrated in a GIS and produce land use maps according to the results of a transportation / land use model. The relevance of this methodology is quite clear because it permits to produce theoretical land use maps with a set of variables representing spatial accumulation organized as rules. The second objective of the paper is to explore the possibility to represents a complex urban system in the form of a structured information pattern. Artificial intelligence, more precisely knowledge based computing, offers methodologies to express the available knowledge in a comprehensive framework.

Functional land use is strictly a qualitative term, it cannot have a continual quantitative value. It is however possible to express land use by a discrete value. We propose that the functional land use of type n for a spatial entity i at time t (expressed as L^n_{it}) is a set of quantitative concepts that represent the level of spatial accumulation of movements, transportation infrastructures, economic activities and population for the same spatial entity and at the

same time (expressed as E_{it}^{k}). So functional land use is defined as:

$$L_{it}^{n} = \left\{ E_{it}^{k} \; \forall \; k \right\} \tag{1}$$

Where k is the number of variables. In the most simple case, the number of variables is equal to 4 ($k=4$), one for each concept of spatial accumulation. Therefore, equation 1 can be developed to become:

$$L_{it}^{n} = \{ T , A , S , P \}_{it} \tag{2}$$

where T_{it} is the interaction level of goods, people and information, A_{it} is the accessibility level of transportation infrastructures, S_{it} is the economic structure and P_{it} is the population of a spatial entity i at time t. The two previous equations define one type of land use for a spatial entity. However, a spatial entity can contain several types of land use forming the global land use. The global land use of a spatial entity L_{it} is defined as:

$$L_{it} = \left\{ L_{it}^{n} \; \forall \; n \in i \right\} \tag{3}$$

The global land use of a spatial entity is the set of land use of all types that are an element of that spatial entity. Equation 3 underlines that the functional land use is similar to a procedure of map overlays commonly used in GIS.

It can be deducted from those equations that land use is a set of variables representing the level of spatial accumulation. The problem is based on finding a methodology to define a functional land use for a spatial entity i at time t from four global concepts of spatial accumulation. An approach is to classify a set of quantitative variables for each spatial entity according to a set of criterias. To reach this goal it is necessary to first, define a membership procedure to identify the functional land use. This methodology uses notions of the fuzzy sets theory and require heuristic rules. Secondly, find the links between the related elements to express their relational structure in a comprehensive framework. Semantic networks offer that possibility. Thirdly, we will apply this methodology to identify the functional land uses of Shanghai. The complexity of the urban system of Shanghai will be used to test that new methodology.

2 Heuristic Identification of Urban Functional Land Use

Conventionally, the identification of functional land use is based on a procedure of classification of spatial entities according to a set of criterias, qualitative or quantitative. In this paper, the spatial entities are the urban districts of Shanghai and the criteria of classification are heuristic. Heuristics involve experiential, judgmental knowledge; the knowledge underlying "expertise"; rules of the thumb, rules of good guessing, that usually achieve desired results but do not guarantee them (Feigenbaum and McCorduck, 1984). This approach seems quite unorthodox when compared with the rigorous tradition of quantitative geography that has intensively used classification techniques like cluster, factorial and discriminant analysis. Without denying the relevance of those statistical techniques, we stress that a knowledge-based approach is quite original and potentially more efficient than conventional techniques because results from statistical classification techniques can be transformed in heuristics rules of classification.

To belong to a type of functional land use L^{n}, a set of variables of a spatial entity must have values included in a same set of membership fields. A membership field $M^{k}(L^{n})$ is a continual numerical space that defines the membership of variable k to a set L^{n} (land use type n):

$$M^{k}(L^{n}) = \left[X^{k}(L^{n}) , Y^{k}(L^{n}) \right] \; \forall \; k \tag{4}$$

where $[X^{k}(L^{n}), Y^{k}(L^{n})]$ are the limits of the membership field. $X^{k}(L^{n})$ is the upper limit and $Y^{k}(L^{n})$ is the lower limit. If all the variables of a spatial entity are included in their membership field to a type of functional land use, then

the spatial entity belongs to this type of functional land use. For instance, if the accessibility level, the economic structure, the population density and the interaction level of a spatial entity are within a certain range of values, then that spatial entity belongs to a certain type of land use. From an heuristic point of view, the membership field is what we *believe* the numerical space that define the membership of a variable to a land use type is. The classification procedure of equation 4 is insufficient because we need a level (or probability) of membership for a spatial entity to a type of functional land use. This level is defined by a membership function $u_i(L^n)$:

$$u_{it}(L^n) - [0,1] \tag{5}$$

This function is based on the fuzzy sets logic (Negoita and Ralescu, 1987) and gives a membership value of a spatial entity i to a type of functional land use L^n between 0 and 1. It is not very probable that a spatial entity can have a membership of 1 to a type of functional land use if we consider that a spatial entity generally contains several types of economic activities. To illustrate this, if we take a surface of one square kilometre in an urban area, we are virtually sure that this surface will contain more than one type of land use even if that same surface as been classified by the authorities as a specific land use zone. There is a probability of existence of several types of land use. The more the level of membership ($u_{it}(L^n)$) tends towards 1, the more the type of functional land use L^n is predominant in a spatial entity i. To identify the functional land uses of a spatial entity, two steps are necessary.

The first step consists in finding the membership field ($M^k(L^n)$, see equation 4) for all the variables k used to identify a functional land use L^n. However, we have stressed that the membership field is not sufficient for that purpose. The fuzzy number offers a more complete prospect. The fuzzy number is a concept similar to the membership field besides that it allows to elaborate a membership function. This represents a partial integration of equation 4 and 5 in the sense that we know what is the membership field and how sure we are of that belief. The fuzzy number indicates where the membership function is optimal without mathematically defining this function. A fuzzy number $N^k(L^n)$ that defines the membership of a variable k to a type of functional land use L^n is expressed by:

$$N^k(L^n) - [\, a^k(L^n) \,,\, X^k(L^n) \,,\, Y^k(L^n) \,] \; \forall \; k \tag{6}$$

where: $a^k(L^n)$ = point where for a variable k a membership function $u_{it}(L^n)$ is equal to 1.0 (optimal).
$X^k(L^n)$ = upper limit relatively to $a^k(L^n)$ where a membership function $u_{it}(L^n)$ is equal to 0.
$Y^k(L^n)$ = lower limit relatively to $a^k(L^n)$ where a membership function $u_{it}(L^n)$ is equal to 0.

For example, a fuzzy number $N^k(L^n)=[3,3,2]$ signifies that a spatial entity must have a value between 1 (3-2) and 6 (3+3) for the variable k to belong to that type of functional land use. The membership level is optimal ($u_{it}(L^n)$ equal to 1; maximum certainty) if the variable has a value of 3. Therefore, the functional land use of a spatial entity i at time t can be identified by this heuristic rule:

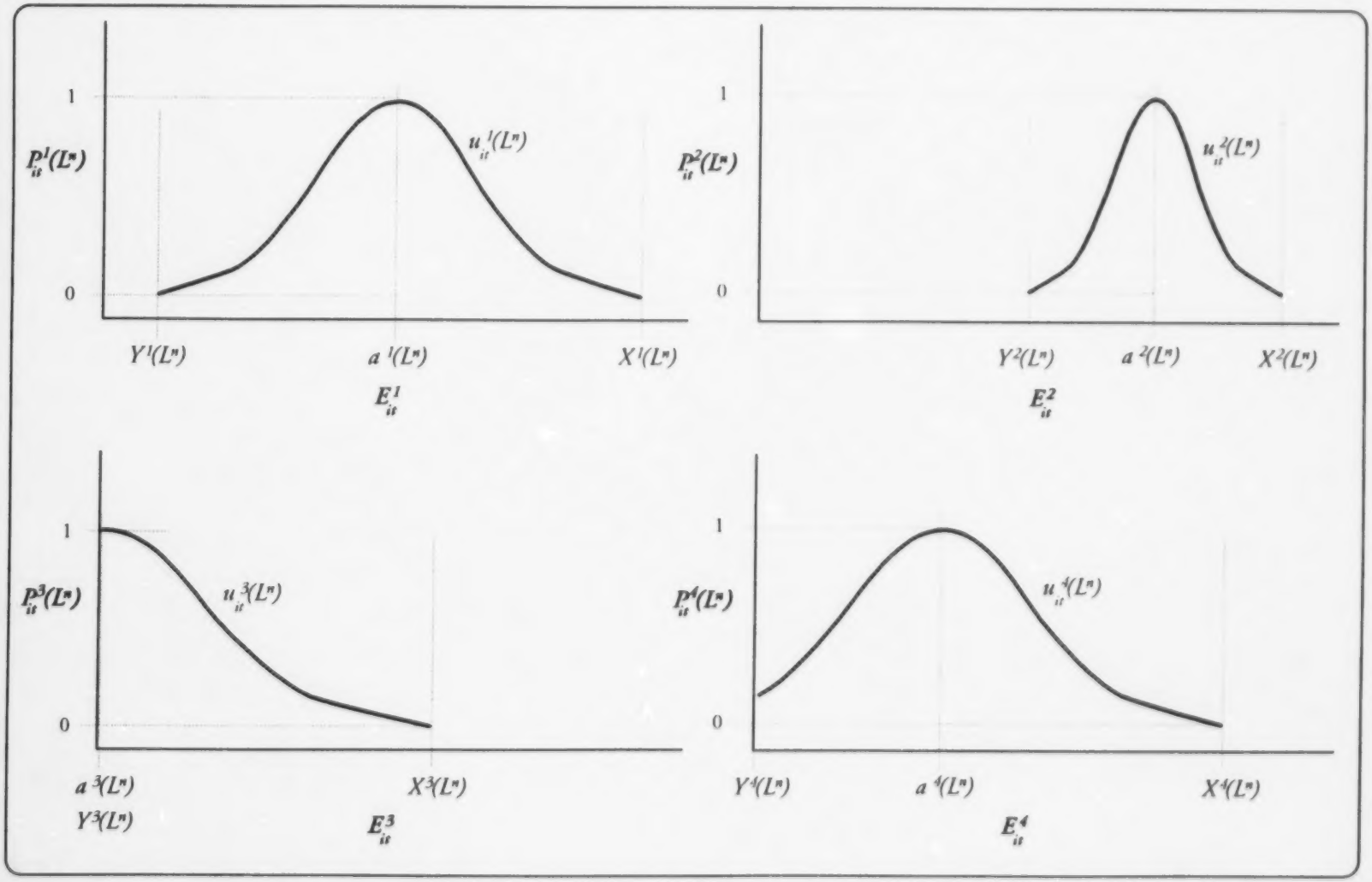
$$i \in L^n \leftarrow \left(E_{it}^k \in N^k(L^n)\right) \; \forall \; k \tag{7}$$

A spatial entity i is a member of a type of land use L^n if and only if all the variables E_{it}^k of that spatial entity are a member of their fuzzy number $N^k(L^n)$. Equation 7 can be developed to become:

$$\exists \; P_{it}(L^n) - [0,1] \leftarrow i \in L^n \leftarrow \left(E_{it}^k \in N^k(L^n)\right) \; \forall \; k \tag{8}$$

There exists a membership value (probability) of a spatial entity i to a type of land use L^n ($P_{it}(L^n)$) if and only if all the variables of that spatial entity are a member of their fuzzy number. This membership value is the result of a membership function that gives results between 0 and 1 ($u_{it}(L^n)$, see equation 5). Figure 2.1 presents the heuristic classification procedure of equation 8 with four variables (E_{it}^1, E_{it}^2, E_{it}^3 and E_{it}^4). If the values of the four variables are within their fuzzy numbers (defined by the $X^k(L^n)$, $Y^k(L^n)$ and $a^k(L^n)$ parameters) then the spatial entity i belongs to the type of land use L^n. The membership functions ($u_{it}^k(L^n)$) are different for each variable. All the membership

Heuristic Classification of Functional Land Use with Fuzzy Numbers **Figure 2.1**



values ($P_{it}^1(L^n)$, $P_{it}^2(L^n)$, $P_{it}^3(L^n)$ and $P_{it}^4(L^n)$) are subject to a weighted sum that gives the global membership value. This is defined in the second step.

The second step consists in evaluating the global membership value of a spatial entity to a type of functional land use defined by equation 8, so $P_{it}(L^n)$ is expressed by:

$$P_{it}(L^n) = \sum_{\forall k} \left(N^k(L^n) \mid u_{it}^k(L^n)\right) \to [0,1]\Big) \, w^k(L^n) \multimap i \in L^n$$
$$\sum_{\forall k} w^k(L^m) = 1 \tag{9}$$

where $w^k(L^n)$ is the weight of variable k on the identification of the land use type L^n. $P_{it}(L^n)$ is the weighted sum of the results of the membership functions to a type of functional land use for each variable k for a spatial entity i at time t. There exists very few empirical evidence on the weight values of several variables characterizing spatial accumulation for the identification of land use. In this article, we propose a simple membership function mathematically defined by $u_{it}(L^n)=f(\sqrt{x})$. No empirical evidence suggests a specific type of membership function, but functions simulating normal-like distributions seam suitable. By integrating equation 9 with our membership function, the global membership value of a spatial entity to a type of functional land use $P_{it}(L^n)$ is defined as:

$$\begin{aligned}
&\exists\, P_{it}(L^n) \multimap i \in L^n \\
&P_{it}(L^n) = \sum_{\forall k} c \sqrt{\left|a^k(L^n) - |a^k(L^n) - E_{it}^k| - |a^k(L^n) - X^k(L^m)|\right|}\, w^k(L^n) \text{ , } \textit{if } E_{it}^k < a^k(L^n) \\
&P_{it}(L^n) = \sum_{\forall k} c \sqrt{\left|a^k(L^n) - |a^k(L^n) - E_{it}^k| - |a^k(L^n) - Y^k(L^n)|\right|}\, w^k(L^n) \text{ , } \textit{if } E_{it}^k > a^k(L^n) \\
&P_{it}(L^n) = 1.0 \text{ , } \textit{if } E_{it}^k = a^k(L^n)
\end{aligned} \tag{10}$$

and where:

$$c = \frac{1}{\sqrt{|a^k(L^n) - |a^k(L^n) - X^k(L^n)||}} \text{ , } \textit{if } E_{it}^k < a^k(L^n)$$
$$c = \frac{1}{\sqrt{|a^k(L^n) - |a^k(L^n) - Y^k(L^n)||}} \text{ , } \textit{if } E_{it}^k > a^k(L^n) \tag{11}$$

where: E_{it}^k = value of a variable k for a spatial entity i.
$a^k(L^n)$ = point where the membership function $f(\sqrt{x})$ is equal to 1.0.
$X^k(L^n)$ = upper limit relatively to $a^k(L^n)$ where a membership function $u_{it}(L^n)$ is equal to 0.
$Y^k(L^n)$ = lower limit relatively to $a^k(L^n)$ where a membership function $u_{it}(L^n)$ is equal to 0.
c = constant of proportionality.

The accuracy of the developed methodology is based on the theoretical knowledge of the relationship between variables of spatial accumulation and land use types. To express those relationships, we have proposed the fuzzy number $N^k(L^n)$ that defines the boundaries inside which a variable have a probability to represent a land use type. A set of fuzzy numbers can be conceived as a set of heuristic rules to identify the functional land uses of a spatial entity. This classification procedure is compatible with the technology of expert systems if we consider fuzzy numbers as heuristic knowledge about functional land use. If we extend that assumption, the operational version of the methodology as a computer program can be conceived as a domain specific expert system. However, as we know, rules are not laws and the result will be reliable as much as the rules are reliable. This put an emphasis on the theoretical knowledge on land use.

The developed methodology offers a way to identify the most probable functional land uses for a spatial entity at a point in time. However, it does not indicate the surface occupied by each type of land use inside a spatial

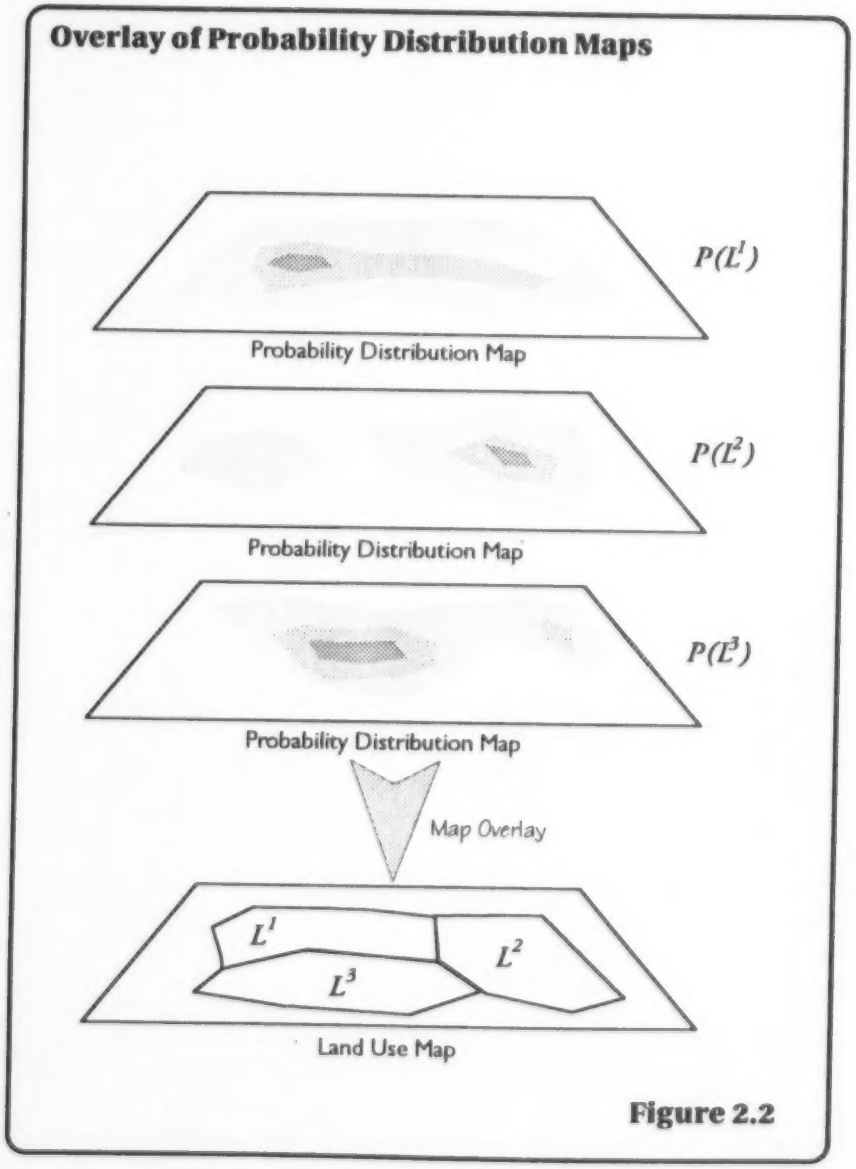

Overlay of Probability Distribution Maps
$P(L^1)$
Probability Distribution Map
$P(L^2)$
Probability Distribution Map
$P(L^3)$
Probability Distribution Map
Map Overlay
L^1
L^2
L^3
Land Use Map

Figure 2.2

entity. We propose that the surface occupied by a type of land use L^n in a spatial entity i (expressed as $S_{it}(L^n)$) is dependent on its global level of membership to that type of land use ($P_{it}(L^n)$). So:

$$S_{it}(L^n) = \left(\frac{P_{it}(L^n)}{\sum_{\forall n} P_{it}(L^n)} \right) D_{it}(L^n)\, s_{it} \qquad (12)$$

where s_{it} is the surface of the spatial entity i at time t and $D_{it}(L^n)$ is the density of land use type L^n. The highest the level of membership of a spatial entity to a land use type, the highest the probability that the type of land use occupy a large portion of the surface of the spatial entity. However the density of land use is often unknown and if $D_{it}(L^n)$ is kept constant, equation 12 will can give unreliable results. A way to bypass that limitation is to identify land use according to its density. For instance, residential land use can be divided in categories like high density residential, low density residential etc... This enable the removal of $D_{it}(L^n)$ from equation 12.

The membership values $P_{it}(L^n)$ can express a spatial probability distribution for each type of land use over an urban area ($P(L^n)$). Figure 2.2 presents the overlay of probability distribution maps. A spatial probability distribution is an extrapolation of membership values assigned to centroids, that is the centre points of each spatial entity. Extrapolation is a procedure available in most GIS that permits to represent a whole statistical surface from a limited number of observed samples. Each type of land use will have its spatial probability distribution map, and through map overlay, the production of land use maps is possible. However, since information is losed on the land use map, it is preferable to use the probability distribution maps instead of their synthesis as a land use map.

3 Land Use as Knowledge about the State of an Urban System

The previous section as looked upon a methodology to identify the functional land use with heuristic rules, which can be conceptualized as acquiring more knowledge on an urban system. If we know the land use types of the spatial entities of an urban system, our knowledge of that system has increased. What we suggest is a relational structure of the information relevant to land use with the purpose to organize it as knowledge, that is the formalization of available information that can be used for inference. The methodology we use to represent land use as knowledge is an adaptation of the *conceptual dependency theory* developed for research in knowledge-based computing (Schank, 1975) in the form of a *semantic network*.

> **Definition 3.1.** A semantic network is a knowledge representation formalism that presents each concept as a node in a graph and the relations between concepts as labelled arcs between nodes; the meaning of a concept is derived from its relationships to all other concepts in the network (Chabris, 1989). The semantic network is a directed acyclic graph. Lets define a semantic network by $S = (\Omega, N, L)$, where Ω is the knowledge domain, N is a set of nodes and L is a set of links between nodes. The knowledge domain is simply the type of information contained in the semantic network.

In our case we use an elaborate semantic structure where a node is called a *frame*. A frame is a data structure, similar to a record in a relational database, which represents a parcel of information with a name and various properties. The links between frames are defined following the value inheritance process. With that process, a frame take all the properties and the values of the previous frames linked to it and therefore allowing to deduce implicit information that is not explicitly represented in the network. For example, if Socrates (frame X) is a man (frame Y) and all man are mortals (frame Z), thus it is logically easy to concludes through inference that Socrates is mortal. This is logically expressed by:

$$X \rightarrow Y \wedge Y \subset Z \therefore X \subset Z$$

A frame can represent an element or a set. For instance, Socrates is an element while man and mortals are sets. The logical links between the frames are qualitative and quantitative, depending on the nature of the knowledge. In our case, three types of links are used; instance, element and property.

Definition 3.2. An instance link, expressed by $\Rightarrow$, shows that one frame is an example of another

frame like *frame[a]* ⇒ *frame[b]*. This type of link is often referred as an ISA link.

Definition 3.3. An element link, expressed by ⊂, is when a frame is a part of another, like *frame[a]* ⊂ *frame[b]*. *frame[a]* is a subset of *frame[b]*, where *frame[a]* can be a set or not.

Definition 3.4. A property link, expressed by (⊸), is when a frame is a characteristic of another, like (*frame[a]* ⊸ *frame[b]*).

The knowledge representation of a semantic network can be conceived by three structural properties: The *knowledge domain*, the *inference network* and the *inference path*.

Definition 3.5. The set of frames and links in a semantic network represents the available knowledge or more precisely the knowledge domain (Ω). Let the knowledge domain Ω = (*F*,*L*), where *F* is the set of all frames in Ω, and *L* is the set of all links between frames in Ω.

Definition 3.6. A goal *G* is a desired inference for a semantic network *S*. Goals can be single or multiple. A single goal gives all the inference paths from one frame. A multiple goal tries the verify an hypothetic relationship between two or more frames.

Definition 3.7. An inference network (Ω,*G*) for a goal *G* in a knowledge domain Ω, is a decomposition of the links in Ω that are related to the goal *G*.

Definition 3.8. An inference path (Ω,*L(G)*) that is a sequence of links that has *G* as a goal (*L(G)*) The sequence of links is find through a value inheritance procedure.

Figure 3.1 describes knowledge based computing with semantic networks that are composed of frames *F* and links *L*. Starting from a semantic network (*F-1*,*L-1*) which have limited knowledge, a knowledge acquisition procedure can be performed by finding links between existing frames and by adding new frames (*F*=*F*+*1*, *L*=*L*+*1*). The result is a new semantic network *(F,L)* that possesses more information than the previous one. Information in both semantic networks can be extracted through query. An Expert System, using a semantic network as an information database can perform two basic procedures. The first, knowledge extraction (a low level inference), use the available information to find implicit relations through value inheritance. This is the available inference structures for knowledge extraction with our three types of links:

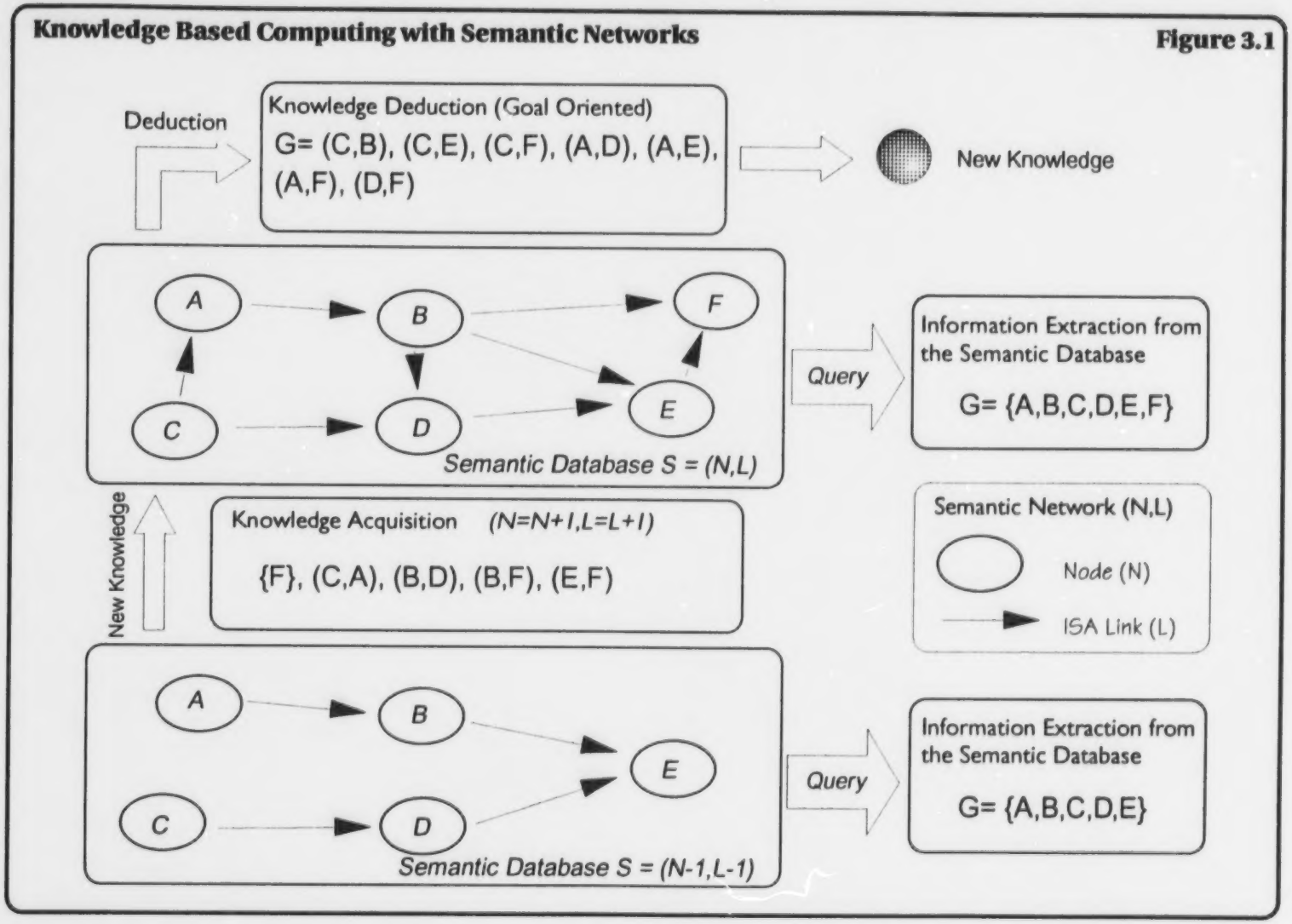
A⇒C :- A⇒B⇒C	A¬⇒C :- A⊂B⇒C	A¬⇒C :- A⊸B⇒C
A⊂C :- A⇒B⊂C	A⊂C :- A⊂B⊂C	A¬⊂C :- A⊸B⊂C
A⊸C :- A⇒B⊸C	A⊸C :- A⊂B⊸C	A⊸C :- A⊸B⊸C

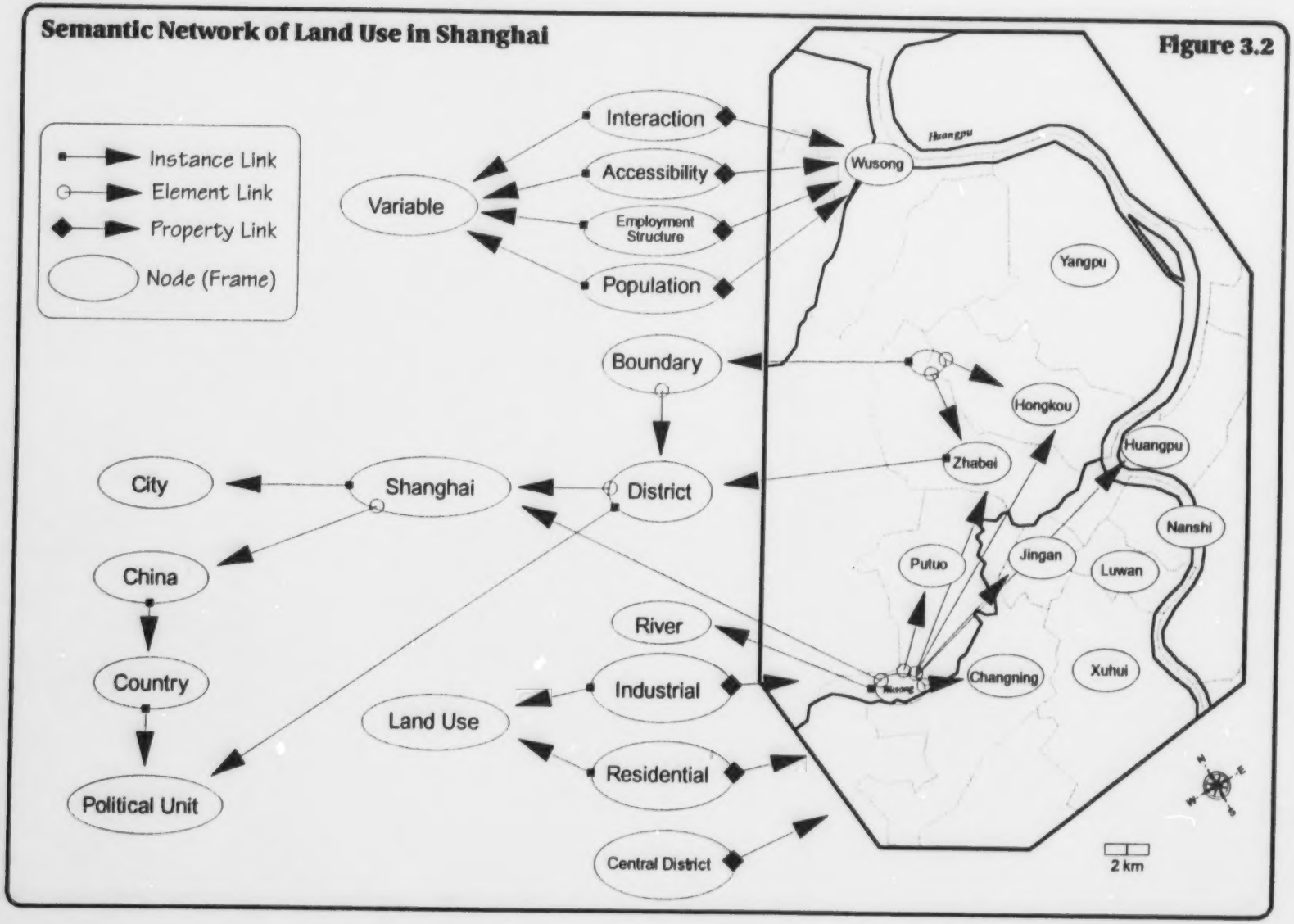
The second, knowledge deduction (an higher order inference), use the information in the semantic network to produce information that answer goals. Figure 3.2 represents a simple frame based semantic network applied to the identification of the functional land use in Shanghai. If the single goal *G*= **Zhabei** (find all the possible paths starting from the frame Zhabei) is given to the semantic network *S*, the inference path is:

(Ω,*L(G)*)= **[Zhabei] ⇒ [District],**
⇒ [Political Unit].
⊂ [Shanghai],
⊂ [China] ⇒ [Country] ⇒ [Political Unit].
⇒ [City].

Semantically speaking, Zhabei is a District in Shanghai, and Shanghai is a city in China that is a country. A district and a country are political units. We can implicitly deduct that Zhabei is in China, but this can be done by the goal *G*= **Zhabei with China**, which will gives as positive answer that Zhabei is in China. The knowledge domain on figure 3.2 is very limited and incomplete, but the objective is here to define a simple knowledge domain of land use. As we see, the property links between land use type frames and districts frames are not defined, because we don't know the relationships between those frames yet. Those links identify to what types of functional land use each urban district of Shanghai belongs. They will be subsequently defined with the heuristic procedure of identification of functional land use previously developed. This procedure will increase the knowledge of the semantic network by adding links between frames.

Knowledge Based Computing with Semantic Networks **Figure 3.1**




Semantic Network of Land Use in Shanghai
Figure 3.2
Instance Link
Element Link
Property Link
Node (Frame)
Variable
Interaction
Accessibility
Employment Structure
Population
Boundary
Shanghai
City
China
Country
Political Unit
District
River
Land Use
Industrial
Residential
Central District
Wusong
Yangpu
Hongkou
Zhabei
Huangpu
Nanshi
Putuo
Jingan
Luwan
Changning
Xuhui
2 km

The advantage of the semantic network theory to the representation of land use as knowledge is that it gives a highly adaptable knowledge structure. The nature of frames and their relationships is not previously fixed. For example, on figure 3.1 land use have only two instance frames: industrial and residential, and there is four types of variables: interaction, accessibility, employment structure and population. If needed, more land use type or variable frames could be added to provide more domain knowledge. This will add more relationships to the whole semantic network and therefore make it more complex.

In order to acquire knowledge on an urban system by identifying the functional land use of a spatial entity, the previous section has stressed the requirement of fuzzy numbers. Considering the available data on the urban districts of Shanghai (see annex), table 3.1 presents the fuzzy numbers ($N^k(L^n)$) and variable weights ($w^k(L^n)$) to classify each urban district to specific types of land use.

Table 3.1: Heuristic classification of Functional Land Use by Fuzzy Numbers

Land Use Type	Input[a]	Output[b]	Heavy[c]	Light[d]	Population[e]
Heavy Industrial	[8,2,7] 0.60[f]		[10,0,7] 0.40		
Light Industrial	[8,2,7] 0.60			[10,0,7] 0.40	
High Density Residential		[5,5,4] 0.25			[10,0,5] 0.75
Medium Density Residential		[4,6,3] 0.25			[4,2,2] 0.75
Low Density Residential		[2,2,1] 0.25			[2,2,1] 0.75
Central District	[10,0,3] 1.00				

[a] Input is the number of persons arriving in a district by transit.
[b] Output is the number of persons departing from a district by transit.
[c] Heavy is the production in Yuan of the heavy industrial sector.
[d] Light is the production in Yuan of the light industrial sector.
[e] Population is the density of persons by square kilometre.
[f] Weight of the variable in the identification of that type of land use.

The fuzzy numbers are defined according to some basic assumptions. Industrial land uses have a high concentration of industrial production and are the destination of high transit movements of labour. Residential land uses have a number of persons departing and a high density of population by square kilometres. Central district is not a land use type but a state of a district that have a high number of persons who travel to it.

4 Application to Shanghai

Shanghai is the most important port and centre of industrial production of China. The city is facing important changes in the nature of its industrial structure. With the growing trade with the international market, the industrial structure is gradually changing from heavy capital intensive industries towards light labour intensive industries. Authorities are looking upon new methodologies to help the planning process of that complex urban system.

The methodology of identification of urban functional land use was implemented as a computer program named FLUI (Functional Land Use Identification). Table 3.1 was transformed in a rule file that the program used to analyze a data file (see annex) in order to produce membership values for specific types of land use for each urban district of Shanghai. Table 4.1 presents the results.

Table 4.1: Membership Values of the Districts of Shanghai to Land Use Types

District	Heavy Industrial	Light Industrial	HD Residential	MD Residential	LD Residential	Central district
Changning				0.3667	0.3381	
Hongkou	0.3160	0.3937	0.2375	0.2345		0.6532
Huangpu			0.2308	0.1451		
Jingan			0.4672			
Luwan			0.4590			
Nanshi	0.1453		0.2151	0.3310		
Putuo		0.3108		0.5276		
Wusong	0.2323				0.3536	
Xuhui	0.2191	0.3820		0.3099	0.3905	
Yangpu	0.4690	0.4690		0.3454	0.3520	1.0000
Zhabei	0.2400			0.4536	0.1331	0.1000

Figure 4.1 illustrates a part of the semantic network of figure 3.2 with established links between land use types and districts. With that process, we have acquired more knowledge about the functional land use of Shanghai. Or course, the knowledge domain is small and highly specific. If we ask the semantic network the query G= **Land Use with Nanshi** the inference path is:

$(\Omega,L(G))$= **[Heavy Industrial]** ⊢ **(0.1453) [Nanshi].**
→ **[Land Use].**
$(\Omega,L(G))$= **[HD Residential]** ⊢ **(0.2151) [Nanshi].**
→ **[Land Use].**
$(\Omega,L(G))$= **[MD Residential]** ⊢ **(0.3331) [Nanshi].**
→ **[Land Use].**

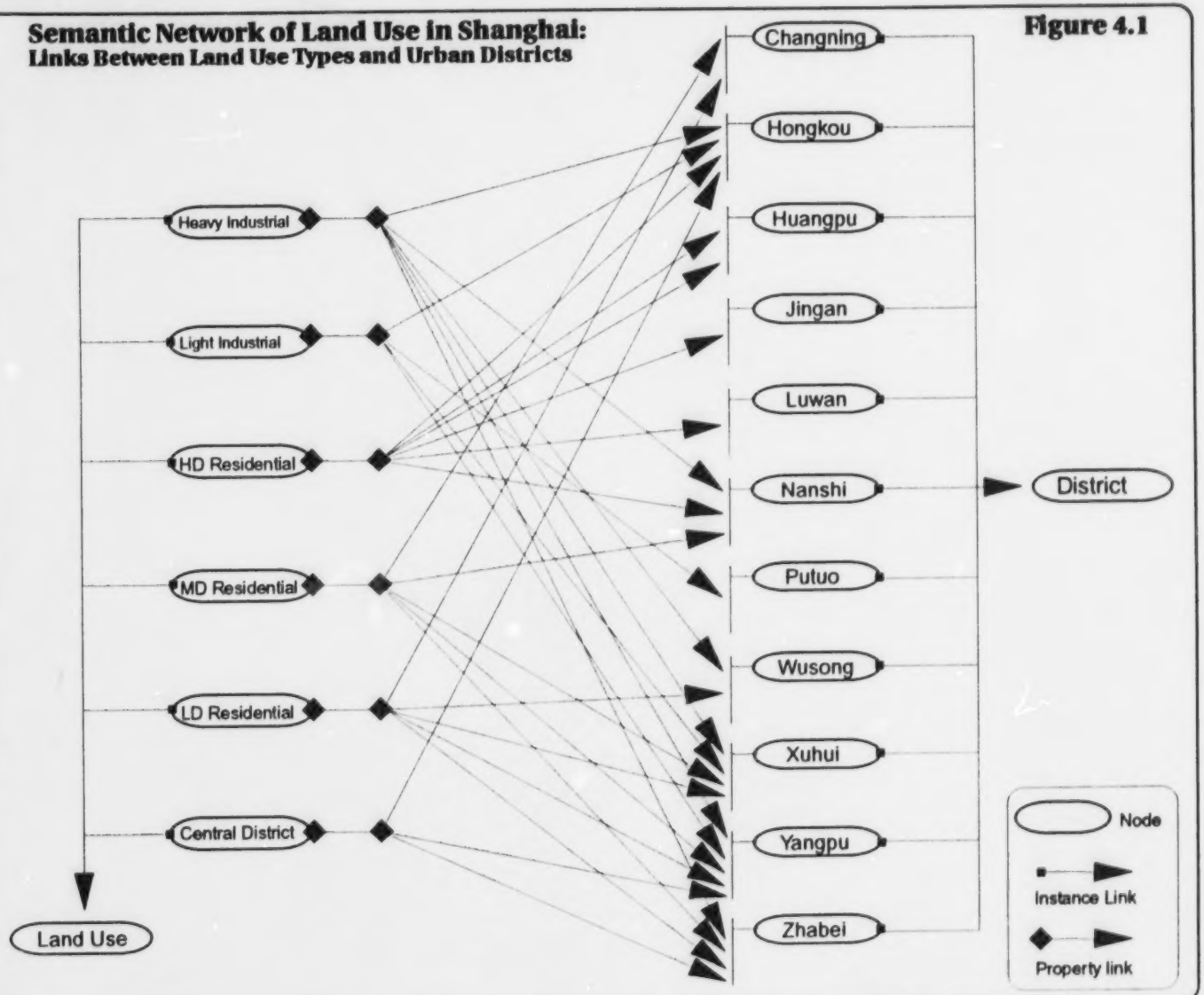
The membership values are generally low. This let suppose that there exists other types of land use in each district and the available data is affected by their presence. For instance, no analysis was undertaken to find membership values for land uses like commercial and institutional due to the lack of available data.

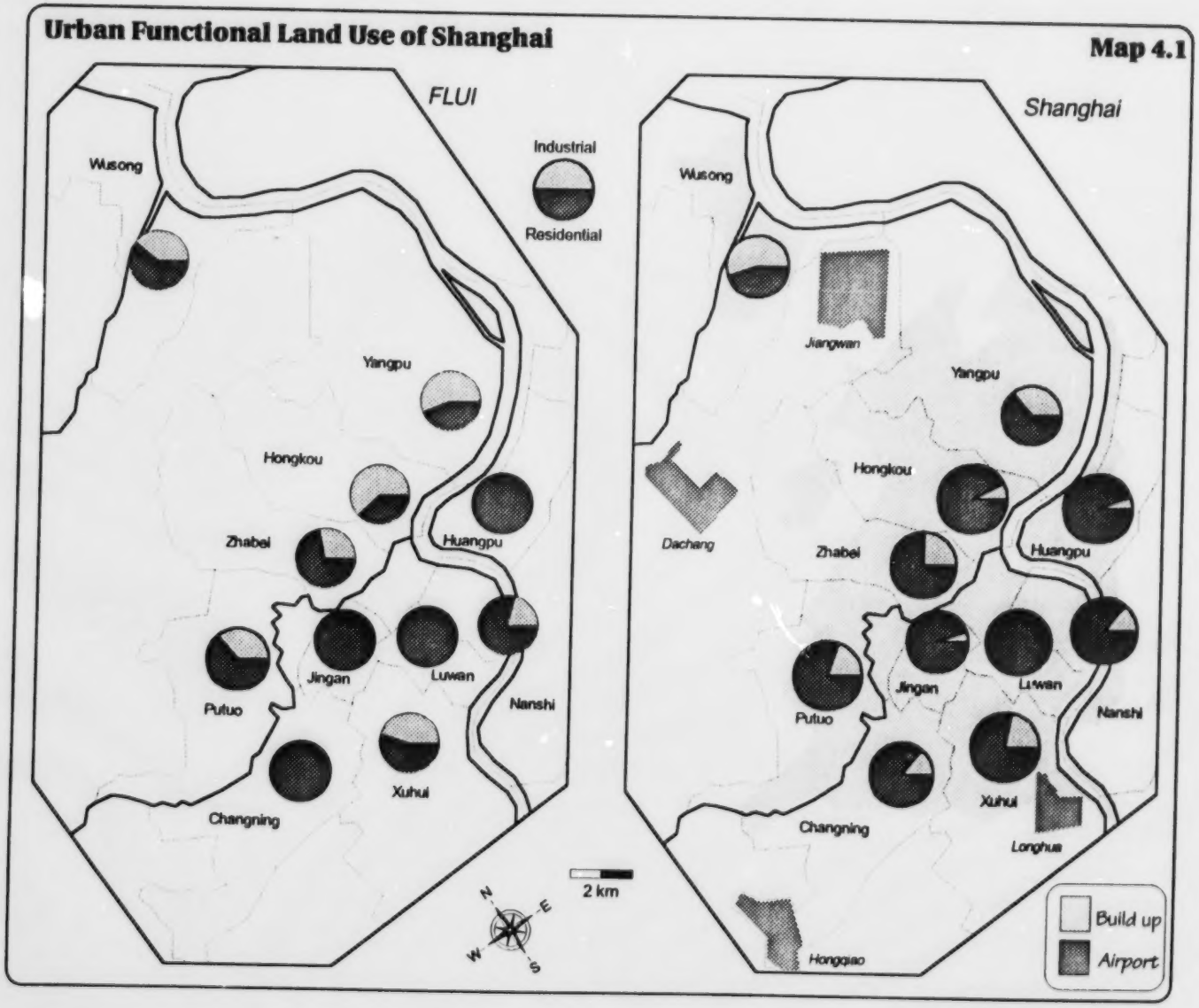
The fact that no district in the central area has been classified with the attribute Central district underline that Shanghai is a polynuclear city. This is quite in discordance with the objective of the authorities to create a central business district in the western part of the Huangpu district (see figure 3.2). Central (highly attractive) districts are Hongkou, Yangpu and Zhabei, all located to the northern part of the city. They correspond to major employment zones in the industrial sector.

Map 4.1 presents a comparison between the land use calculated by FLUI and the real land use of Shanghai. The results are expressed as the percentage of the build up area occupied by each type of land use using equation 12 for the FLUI program, and using the real surface occupied by land use for Shanghai. Since the available data has permitted the calculation of membership values for only industrial and residential land uses, the comparison is restricted to those land uses. Most of the results are very similar, underlining the accuracy of the rules. However, results do not concord in the cases of Xuhui, Hongkou and for some extent Yangpu districts. The main factor behind this is the data about the industrial output says nothing about the density of industrial land use. For instance, the main difference between the calculated land use surface and the real one for those districts could be attributed to the fact that those districts have an important part of their industrial output in the light industrial sector. In Shanghai,

Semantic Network of Land Use in Shanghai:
Links Between Land Use Types and Urban Districts

Figure 4.1




Urban Functional Land Use of Shanghai
Map 4.1
FLUI
Industrial
Residential
Wusong
Yangpu
Hongkou
Zhabei
Huangpu
Jingan
Luwan
Putuo
Nanshi
Xuhui
Changning
2 km
N
E
W
S
Shanghai
Jiangwan
Dachang
Longhua
Hongqiao
Build up
Airport

this sector is characterized by an high density several storeys industrial infrastructures. Therefore the procedure attribute them more space than they are entitled to. The Wusong district shows opposite results. The real surface occupied by the industrial land use is more important than the one found by the FLUI program. This district is characterized by space consuming heavy industrial infrastructures, notably steel plants. Rules for the identification of industrial land use and several density related urban land uses should desegregate land use types by density if no density information is available, which was our case.

5 Conclusion

Contemporary urban areas present an intense mixture of several types of land use including non urban and are extending over a vast territory. Over the past, the land use for a spatial entity was arbitrary classified following a set of criterias. However, the membership of a spatial entity to a land use type is not absolute, but must rather be viewed as the overlapping of diverse types of transportation infrastructures, movements, economic activities and population, each having its own type of land use. Considering this, a methodology to identify the functional land uses of spatial entities was developed. The approach was based on the assumption that a spatial entity has a membership value (or probability) to be associated with a type of land use. The assignation of a membership value was done with an heuristic classification procedure using elements of fuzzy logic. We have advocated that land use can be conceived as knowledge about an urban system, and developed a simple semantic network that represents a knowledge domain. The methodology help the semantic network "learn" (knowledge acquisition) about an urban system by finding the links between spatial entities and types of land use.

Knowledge representation is a part of the available knowledge-based computing techniques that deserve some considerations because they offers a way to formally express the knowledge in several geographical domains were knowledge is quite scattered. The approach of geographical knowledge engineering is challenging in the way that we must reconsider the way we think about a problem and what we believe is relevant to it. As always in geography and in all other fields were judgment and experience play a fundamental role, rules are not laws and the knowledge is always not entirely indubitable. Artificial intelligence uses mathematical tools like the fuzzy logic and probabilistic reasoning (see Neapolitan, 1990 and Negoita and Ralescu, 1987) to handle that constraint.

Bibliography

Berechman, J. and K.A. Small (1988) "Research policy and review 25. Modeling land use and transporation: an interpretive review for growth areas", Environment and Planning A, Vol. 20, pp. 1285-1309.

Chabris, C.F. Artificial Intelligence and Turbo C, Dow Jones-Irwin, Homewood, IL.

Feiganbaum, E.A. and P. McCorduck (1984) The Fifth Genaration: Artificial Intelligence and Japan's Computer Challenge to the World, The New American Library of Canada, Scareborough, Ontario.

Grothe, M. and M.J. Scholten (1990) "Evaluation of the application of network analysis: Theory versus practice", EGIS '90, Proceedings, First European Conference on Geographical Information Systems, Armsterdam, The Netherlands, Vol. 1, pp. 410-417.

Lindgren, D.T. (1985) Land Use Planning and Remote Sensing, M. Nijhoff, Hingham, MA.

Mather, A.S. (1986) Land Use, Longman, New York.

Neapolitan, R.E. (1990) Probabilistic Reasoning in Expert Systems: Theory and Algorithms, John Wiley & Sons, New York.

Negoita, C.V. and D. Ralescu (1987) Simulation, Knowledge-Based Computing, and Fuzzy Statistics, Van Nostrand Reinhold, New York.

Schank, R.C. (1975) Conceptual Information Processing, North Holland, Amsterdam.

Webster, F.V. and P.H. Bly (1987) "Changing Pattern of urban travel and implications for land use and transport strategy", Transportation Research Record 1125, Washington, D.C., pp. 21-28.

van Est, J.P. and C.M. Sliepen (1990) "Geographic information systems as a basis for interaction modelling: an application", EGIS '90, Proceedings, First European Conference on Geographical Information Systems, Armsterdam, The Netherlands, Vol. 1, pp. 298-308.

Annex

Table: Some Geographical Variables on the Urban Districts of Shanghai (except Minhang), 1987.

i	T_{si}	T_{is}	P_i	S_i^1	S_i^2	A_i
Changning	3784.1	6460.2	17227	27.86	19.89	10244.3
Hongkou	11475.2	5850.2	35901	54.21	32.73	17325.4
Huangpu	2367.2	3904.8	36562	18.66	10.66	6272.0
Jingan	3511.2	5224.3	65591	29.61	11.53	8735.5
Luwan	3590.4	5294.0	63973	20.67	12.86	8884.4
Nanshi	4494.4	8276.2	33542	22.64	25.98	12770.6
Putuo	8149.9	20337.0	22437	44.35	19.32	28486.9
Wusong	2828.9	1650.7	3407	6.93	43.87	4479.6
Xuhui	9195.4	5856.7	15104	56.00	25.80	15052.1
Yangpu	13630.2	4631.1	18378	83.51	74.28	18261.3
Zhabei	9917.9	5459.7	23980	25.02	26.43	15377.6

T_{si} are the inputs of persons by transit of a district (x10 persons).
T_{is} are the outputs of persons by transit of a district (x10 persons).
P_i is the density of population per square kilometre of a district.
S_i^1 is the value of the light industrial production of a district (x100 millions of yuan).
S_i^2 is the value of the heavy industrial production of a district (x100 millions of yuan).
A_i is the accessibility of a district [3].
Sources: Statistical Yearbook of Shanghai, 1988, and O-D matrix of Shanghai, 1982.

[3] $A_i = T_{si} + T_{is}$